It is claimed:

1. A method for retrieving answers to questions from an information retrieval system comprising:

generating a set of phrases that identify different categories of questions;

generating candidate transforms for each phrase;

weighting the candidate transforms;

ranking the candidate transforms;

applying the transforms to an information retrieval system.

2. A method as in claim 1 further comprising:

filtering candidate transforms prior to weighting.

3. A method as in claim 2 further wherein:

natural language processing techniques are used for filtering.

4. A method as in claim 3 wherein:

a natural processing technique used is part-of-speech tagging.

5. A method as in claim 4 wherein:

Brill's part-of-speech tagger is used.

6. A method as in claim 3 wherein:

the natural language processing techniques used are feature selection techniques.

7. A method as is claim 1 wherein:

the questions are categorized by similar goals.

8. A method as in claim 7 wherein:

the categories are identified using an *n*-gram approach.

9. A method as in claim 8 wherein:

phrases are generated by computing the frequency of all *n*-grams of length *minQtokens* to *maxQtokens* words, with all *n*-grams anchored at the beginning of the questions.

10. A method as in claim 9 wherein:

all *n*-grams that occur at least *minQphraseCount* times are used for generating question phrases.

11. A method as in claim 7 wherein:

the input to generating the set of phrases is a set of questions.

12. A method as in claim 7 wherein:

the output to generating the set of phrases is a set of question phrases that can be used to classify questions into respective question types.

13. A method as in claim 1 further comprising:

filtering the generated phrases.

14. A method as in claim 13 wherein:

natural language processing techniques are used for filtering.

15. A method as in claim 14 wherein:

a natural processing technique used is part-of-speech tagging.

16. A method as in claim 15 wherein:

Brill's part-of-speech tagger is used.

17. A method as in claim 14 wherein:

the natural language processing techniques used are feature selection techniques.

18. A method as in claim 1 wherein:

generating candidate transforms comprises generating initial candidate transform phrases.

19. A method as in claim 18 wherein:

initial candidate transforms are generated by using a collection of *Question/Answer* pairs.

20. A method as in claim 19 further comprising:

filtering initial candidate transform phrases.

21. A method as in claim 20 wherein:

initial candidate transforms are filtered by minimum co-occurrence.

22. A method as in claim 20 further comprising:

weighting filtered initial candidate transforms.

23. A method as in claim 22 further comprising:

filtering all weighted initial candidate transforms.

24. A method as in claim 19 wherein:

the collection of pairs has been tagged with a part-of-speech tagger.

25. A method as in claim 24 wherein:

Brill's part-of-speech tagger is used.

26. A method as in claim 19 wherein:

initial candidate transform phrases are filtered by eliminating generated answer phrases that contain a noun.

27. A method as in claim 19 wherein:

all potential answer phrases are generated from all of the words in the prefix of *Answer* for each *Question/Answer* pair where a prefix of *Question* matches each question phrase.

28. A method as in claim 27 wherein:

n-grams of length *minAtokens* to *maxAtokens* words are used, starting at every word boundary in the first *maxLen* bytes of the *Answer* text.

29. A method as in claim 28 wherein:

from the resulting *n*-grams, the *topKphrases* with the highest frequency counts are kept.

30. A method as in claim 19 wherein:

information retrieval techniques for term weighting is applied to rank the initial candidate transforms.

31. A method as in claim 30 wherein:

a Sparck Jones inverse collection frequency weighting scheme that uses relevance information is applied.

32. A method as in claim 30 wherein:

candidate transforms are weighted by assigning to each phrase an Robertson/Sparck Jones term weight with respect to a specific question type.

33. A method as in claim 30 wherein:

the weight is computed for each candidate transform tr_i by computing the count of *Question/Answer* pairs where tr_i appears in the *Answer* to a question matching a question phrase as the number of relevant documents;

considering the number of remaining *Question/Answer* pairs where tr_i appears in the *Answer* as non-relevant, and;

applying the formula $w_i^{(1)} = \dfrac{(r+0.5)/(R-r+0.5)}{(n-r0.5)/(N-n-R+r+0.5)}$.

34. A method as in claim 30 wherein:

term selection weights are computed for each candidate transform.

35. A method as in claim 34 wherein:

term selection weights, wtr_i, for each candidate transform tr_i, are computed as :

$$wtr_i = qtf_i \cdot w_i^{(1)}$$

where qtf_i is the co-occurrence count of tr_i with QP, and $w_i^{(1)}$ is the relevance-based term weight of tr_i computed with respect to QP.

36. A method as in claim 19 further comprising:

sorting the initial candidate transforms into buckets according to the number of words in the transform phrase, and up to *maxBucket* transforms, with the highest values of term selection weights kept from each bucket.

37. A method as in claim 36 further comprising:

filtering and weighting the initial candidate transform prior to sorting.

38. A method as in claim 1 wherein:

ranking the candidate transforms comprises retrieving a set of *Question/Answer* pairs and for each pair and the candidate transforms, applying a transform to each *Question.*

39. A method as in claim 38 further comprising:

sorting *Question/Answer* pairs by increasing answer length prior to ranking the candidate transforms.

40. A method as in claim 38 wherein:

the transforms are encoded so that they are treated as phrases by the information retrieval system.

41. A method as in claim 38 wherein:

a *Question* requires parts of the query in matching pages.

42. A method as in claim 38 wherein:

a Question does not require parts of the query in matching pages.

43. A method as in claim 38 wherein:

multiple transformations are combined into a single query.

44. A method retrieving documents from an information retrieval system comprising:

categorizing questions asked of the information retrieval system into different types;

generating phrases that identify the question types;

generating candidate query transformations for each phrase from a training set of question/answer pairs;

evaluating the candidate transforms on the target information retrieval systems, and;

applying transformations to queries submitted to the information retrieval system.

45. A method for retrieving documents as in claim 44 wherein:

the questions are categorized by similar goals.

46. A method for retrieving documents as in claim 44 wherein:

phrases are generated by computing the frequency of all *n*-grams of length *minQtokens* to *maxQtokens* words, with all *n*-grams anchored at the beginning of the questions.

47. A method for retrieving documents as in claim 46 wherein:

all *n*-grams that occur at least *minQphraseCount* times are used for generating candidate transforms.

48. A method for retrieving documents as in claim 44 further comprising:

filtering the generated phrases.

49. A method for retrieving documents as in claim 48 wherein:

natural language processing techniques are used for filtering.

50. A method for retrieving documents as in claim 49 wherein:

a natural processing technique used is part-of-speech tagging.

51. A method for retrieving documents as in claim 50 wherein:

Brill's part-of-speech tagger is used.

52. A method for retrieving documents as in claim 49 wherein:

the natural language processing techniques used are feature selection techniques.

53. A method for retrieving documents as in claim 44 further comprising:

filtering, weighting, and ranking the candidate query transformations prior to evaluating on the information retrieval systems.

54. A method for retrieving documents as in claim 53 wherein:

natural language processing techniques are used for filtering.

55. A method for retrieving documents as in claim 54 further comprising:

initial candidate transforms are filtered by minimum co-occurrence.

56. A method for retrieving documents as in claim 53 wherein:

generating candidate transforms comprises generating initial candidate transform phrases.

57. A method for retrieving documents as in claim 44 wherein:

filtering initial candidate transform phrases.

58. A method for retrieving documents as in claim 44 wherein:

the training set of pairs are tagged with a part of speech tagger.

59. A method for retrieving documents as in claim 44 wherein:

candidate transforms are filtered by eliminating phrases with nouns.

60. A method for retrieving documents from an information retrieval system comprising:

entering a question whose answer is desired;

classifying the question by matching it with predetermined question phrases;

retrieving the associated question phrases;

rewriting the question by applying each associated question phrase to the question to create transformed queries;

submitting the transformed queries to an information retrieval system;

analyzing the returned documents;

scoring the returned documents;

ranking the returned documents by their respective scores, and;

documents ranked above a predetermined level as the resulting retrieved documents.